

Supplementary Material for:

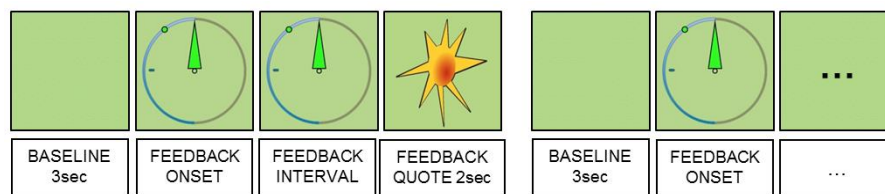
Better than sham? – A double-blind placebo-controlled neurofeedback study in
primary insomnia

Manuel Schabus^{a,b}, Hermann Griessenberger^a, Maria-Teresa Gnjezda^a, Dominik P.J. Heib^{a,b},
Malgorzata Wislowska^{a,b}, & Kerstin Hoedlmoser^{a,b}

^aLaboratory for Sleep, Cognition and Consciousness Research, Department of Psychology,
University of Salzburg, Austria, ^bCentre for Cognitive Neuroscience Salzburg (CCNS),
University of Salzburg, Austria

SUPPLEMENTARY MATERIAL

In the following we illustrate the exact procedure of a single NFT/PFT trial. Several of these trials constituted a block (5min), the exact number of trials depended on how long it took participants to reach the reward criterion; yet usually comprised 13-25 trials. Eight blocks (à 5min) constituted one session or one NFT/PFT training day.



Supplementary Figure 1. Schematic representation of one NFT training trial. For each trial a 3sec (artefact-free) baseline was recorded (represented by the green dot in the circle). Thereafter, participants got immediate feedback about the amplitude in a specific frequency band with a green compass needle. If the threshold (green dot) was exceeded for more than 250ms, a “feedback quote” (reward) was given and a new trial started. Artefacts (blinking, clenching one’s teeth, etc.) terminated the trial and the next trial was started with a new baseline measurement. Figure reproduced from Hoedlmoser et al. (2008).

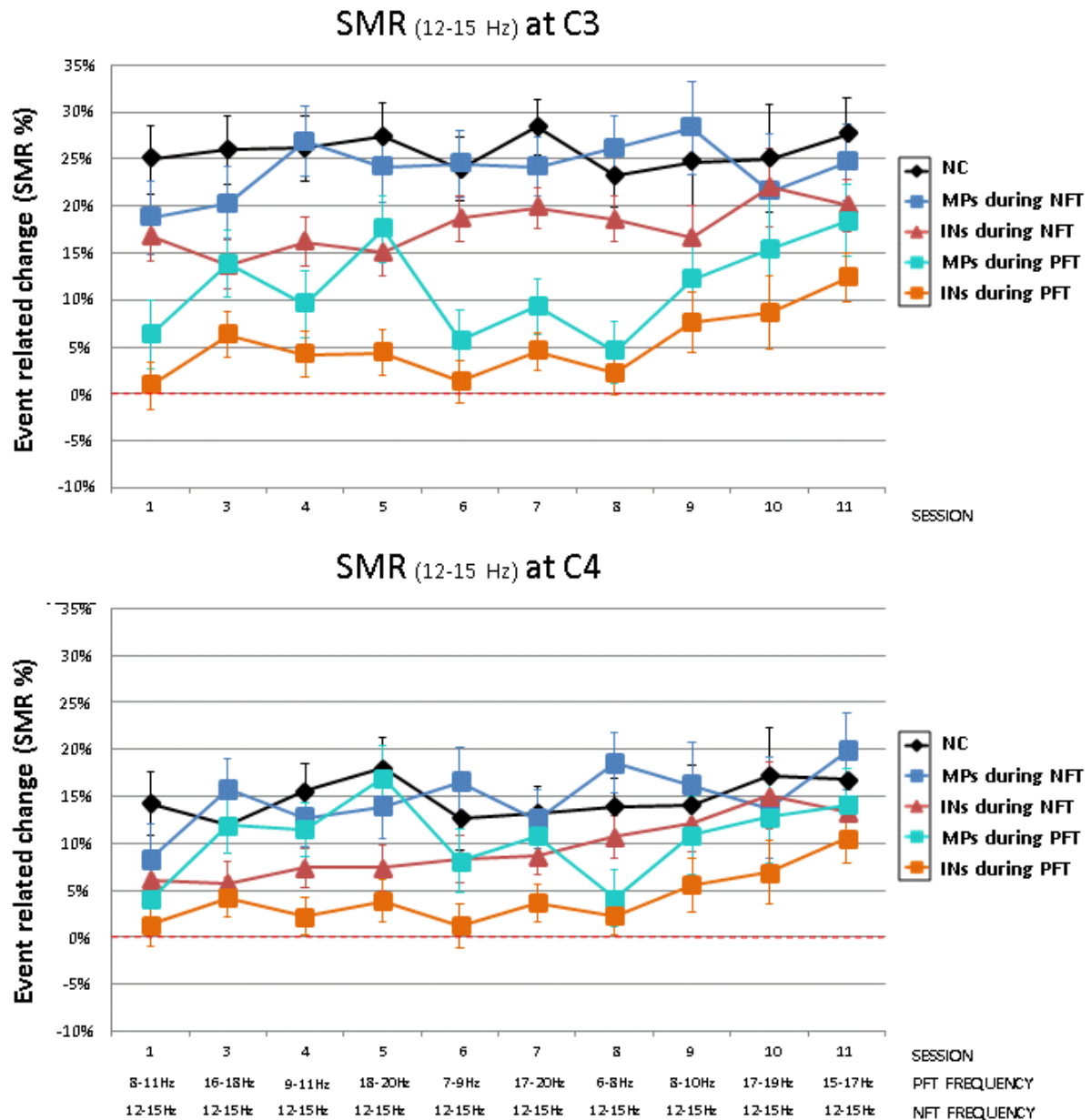
Analysis of the power in frequency bands during Neurofeedback or Placebo-feedback stimulation in insomniacs, misperception insomniacs and neurofeedback controls

To investigate effects during training, we computed an ANOVA and tested for changes in EEG power at electrode C3 and C4 (% change to baseline) during NFT and PFT training as compared to the 3sec baseline preceding each trial. This ANOVA included the factors TIME (10 sessions, due to technical problems during the recording we were unable to use data from sessions 2 and 12), FREQUENCY (Theta, SMR, Beta) and the between-subject factor GROUP (IN during NFT, IN during PFT, MP during NFT, MP during PFT, and healthy NC during NFT)

SMR at C3: The ANOVA (TIME x FREQUENCY x GROUPS) revealed significant main effects for the factors TIME ($F_{9,450} = 6.486$; $p < .001$; $\eta_p = 0.115$) and FREQUENCY ($F_{2,100} = 112.272$; $p < .001$; $\eta_p = 0.692$). However, no significant between groups effect was found when we categorized individuals in five groups (cf. Supplementary Fig.2) ($F_{4,50} = 1.875$; $p = .129$; $\eta_p = 0.130$). Post hoc t-tests of the group differences indicated that effects were generally limited to the SMR frequency band.

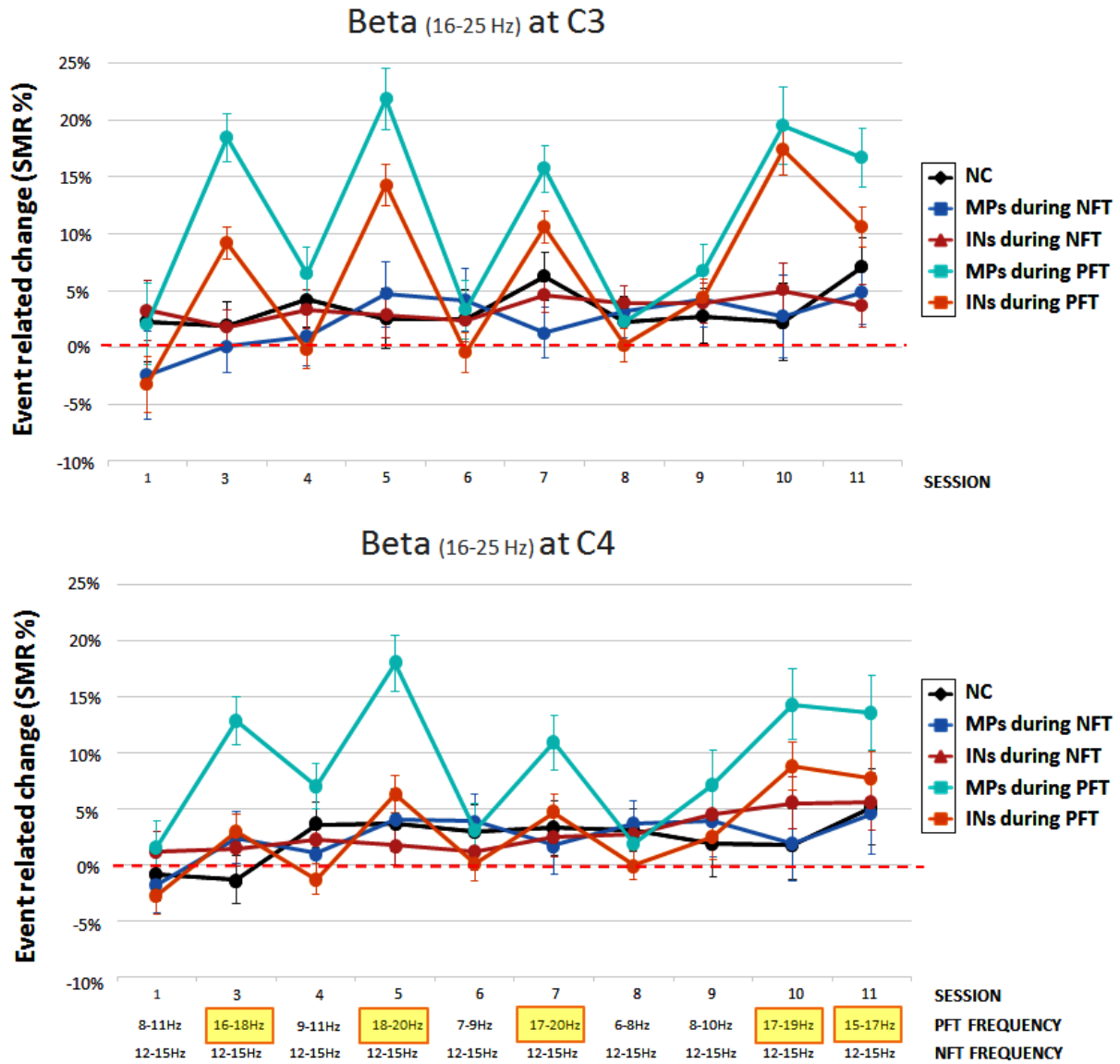
Post hoc t-tests indicated that healthy controls (NC) (doing SMR-NFT) ($M = 25.86$; $SD = 7.04$) exhibited significantly higher SMR values than INs ($M = 5.57$; $SD = 8.68$; $p < .001$) and MPs only during PFT sessions ($M = 11.65$; $SD = 5.46$; $p = .002$). Likewise, MPs during NFT ($M = 24.29$; $SD = 7.56$) outperformed INs ($p < .001$) and MPs during PFT training ($p = .013$). INs training NFT ($M = 18.04$; $SD = 9.68$) were only showing higher SMR values than when training in the PFT condition ($M = 5.57$; $SD = 8.68$; $p < .001$) but not outperforming any other group.

SMR at C4: An analysis of variance (ANOVA) with the dependent variable power change at the contralateral site C4 revealed that there was a significant main effect of TIME ($F_{9,450} = 6.908$ $p < .001$; $\eta_p^2 = .121$) indicating a general increase in SMR power. Furthermore, there was a significant effect of FREQUENCY ($F_{2,100} = 35.468$ $p < .001$; $\eta_p^2 = .415$), but no further significant group differences were found. The interaction between FREQUENCY x GROUPs was highly significant ($F_{8,100} = 7.510$; $p < .001$; $\eta_p^2 = .375$). Post hoc t-tests revealed that neurofeedback controls (NC) ($M = 13.91$; $SD = 6.06$) and misperception insomniacs (MP) in the NFT block ($M = 14.67$; $SD = 6.42$) generally outperformed the insomniac PFT group ($M = 4.32$; $SD = 7.68$). For the theta frequency band no significant group interaction was found. In order to illustrate the topographical specificity of NFT training (on C3) we plot SMR power at electrodes C3 and C4 below (Supplementary Fig. 2).



Supplementary Figure 2. SMR band power during NFT/PFT training blocks. Note the SMR enhancement during NFT sessions in healthy individuals (NC) and insomnia misperception subjects (MPs) vs. “real” insomnia patients (INs) and subjects in the randomized-frequency (placebo-feedback, PFT) protocol. As illustrated, the enhancement of SMR power during NFT as compared to PFT seems to be more pronounced at the site of training (C3). Please note that healthy neurofeedback controls did only take part in the NFT but not PFT training sessions. Abbreviations: NC, healthy neurofeedback controls; MPs, misperception insomniacs; INs, insomniacs.

Theta and beta frequency did not change during NFT training, yet it is obvious that training beta frequency bands (session 3: 16-18Hz, session 5: 18-20Hz; session 7: 17-20Hz; and session 10: 17-19Hz) in the PFT condition also led to higher overall beta during the respective sessions therefore supporting the conclusion that the neurofeedback training worked in MP as well as IN patients (cf. Supplementary Fig. 3). See also statistical details in the following.



Supplementary Figure 3. Beta band power during NFT/PFT training blocks. Note that also beta power increased in those blocks in which beta frequency bands were rewarded (PFT, session 3, 5, 7, 10-11). Again, insomnia misperception subjects were better at learning to enhance the rewarded frequency band as compared to insomnia patients. Please note that neurofeedback controls (NC) did only take part in the NFT but not in the PFT training sessions.

Beta at C3: Specifically, another ANOVA with the factors TIME x GROUP (MP, IN) x FEEDBACK (NFT/PFT) revealed a significant main effect for the factors TIME ($F_{9,153} = 16.069$; $p < .001$; $\eta_p^2 = .486$) and FEEDBACK ($F_{1,17} = 17.329$; $p = .001$; $\eta_p^2 = .505$). Beta-power was higher for PFT ($M = 10.061$; $SD = 1.37$) than for NFT ($M = 2.668$; $SD = 1.25$). Analyses revealed no further significant effects. Interestingly, an additional significant interaction between TIME x FEEDBACK proved to be significant ($F_{9,153} = 14.531$; $p < .001$; $\eta_p^2 = .461$). Post hoc test revealed that this interaction effect was only evident for PFT Feedback ($F_{9,198} = 48.075$; $p < .001$; $\eta_p^2 = .686$) which is clearly related to the fact that during five of the PFT sessions frequencies from the beta range were rewarded. For illustrative

purposes, we also show beta power at electrode C4 (Supplementary Fig. 3, bottom). As for SMR power (cf. Supplementary Fig. 2) the power increase is more pronounced at the left and thus in the rewarded hemisphere (electrode C3) during training.

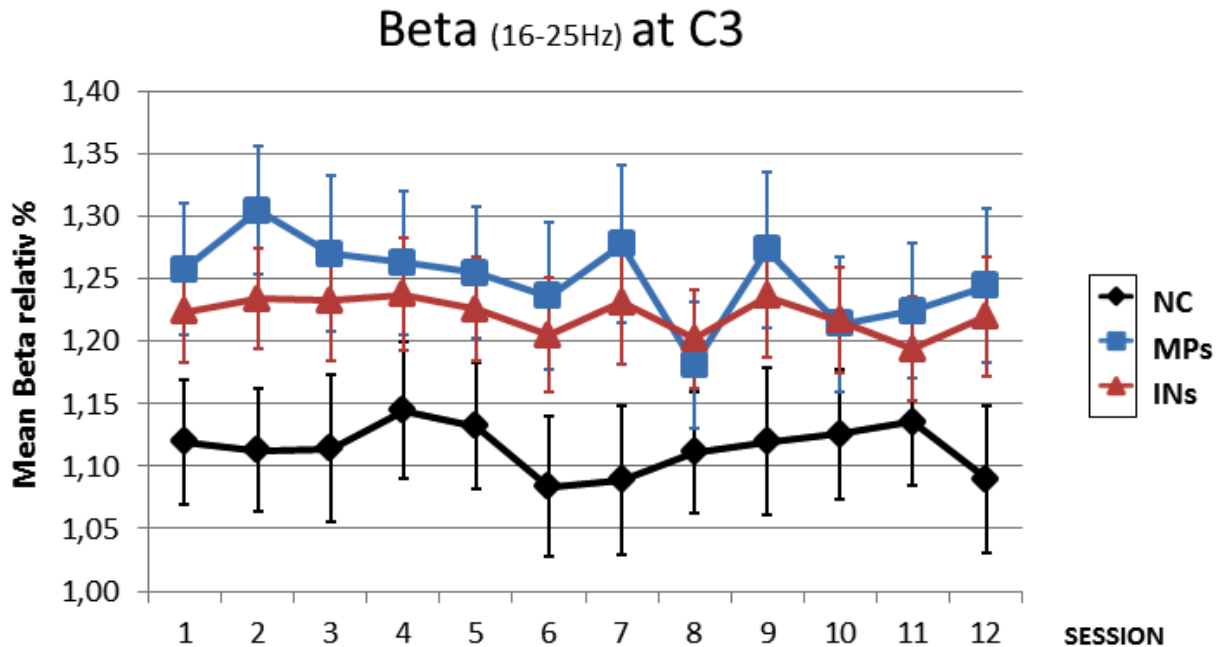
The study therefore revealed a generalization effect with elevation of rewarded frequencies (SMR in NFT, Beta in some PFT sessions) also being observable at the site contralateral (C4) to the rewarded brain signal. This transfer to the other hemisphere appears to be of smaller size but still clearly measurable. In this context it is also interesting to note that the observed frequency response of the EEG is not necessarily specific to the trained frequency band but smears to neighboring frequencies. Further studies that intend applying well-controlled NFT protocols may thus consider adopting inhibition filters for frequencies around the trained frequency band (here, around 12-15Hz). At present we chose not to do so as likely the difficulty for getting rewarded for enhancing a specific frequency band will increase and as the neurofeedback literature is highly ambiguous on optimal training protocols.

Short-Term EEG Effects of NFT/PFT – Beta frequency range

An ANOVA with the within-subject factors FEEDBACK (NFT vs. PFT), ELECTRODE (C3 vs. C4), TIME (12 training blocks) and PRE/POST (before and after each training session) revealed no main effects or interactions with FEEDBACK, or PRE/POST for patients (INs, MPs pooled). However, a highly significant main effect for the factor ELECTRODE was evident ($F_{1, 22} = 12.341$, $p = .002$; $\eta^2 = .359$) with values on C4 ($p = .002$; $M = 1.266$; $SD = .038$) being higher than on C3 ($p = .002$; $M = 1.234$; $SD = .037$).

Since beta has previously been associated with hyperarousal and as there may exist potential group differences (higher beta in patients suffering from insomnia; cf. Perlis et al., 2001) as well as training changes we also focused on beta before and after NFT training blocks in an ANOVA with the factors TIME x PRE/POST NFT x GROUP and the three groups (NC, MP, IN). For the dependent variable relative beta amplitude (i.e., 16-25Hz divided by total-amplitude in the 1-30Hz range) only a marginal significant main effect for TIME was evident ($F_{11, 314} = 1.905$, $p = .070$; $\eta^2 = .58$), yet no further effects were revealed. Importantly, thus also our three groups were not different with respect to (resting) beta amplitude ($F_{2, 31} = 2.035$, $p = .148$; $\eta^2 = .12$) (cf. Supplementary Fig. 4).

In these lines, and given the pioneering work of Peter Hauri (Hauri, 1981; Hauri *et al.*, 1982) we also tried to differentiate anxious and tense insomnia patients from those more relaxed at intake but could not verify that (exclusively) latter are the ones benefiting from SMR-NFT (as evidenced by objective measures of sleep quality).

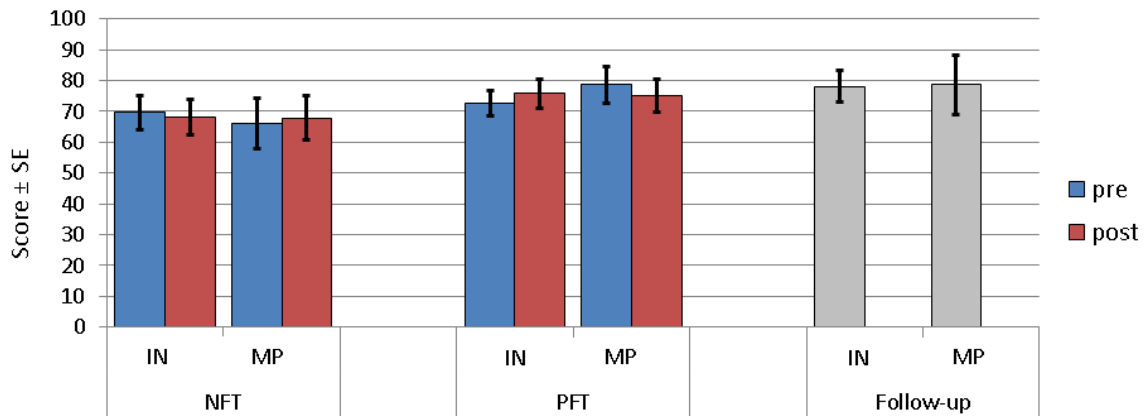


Supplementary Figure 4. Beta band amplitude during rest preceding/following NFT training blocks. A resting condition (with eyes open; electrode C3) directly preceding and following the NFT blocks revealed that beta amplitude did not change through NFT training and that also the three groups did not differ in their baseline beta amplitude. Consequently, beta amplitude is pooled for blocks preceding and following the NFT training blocks. Note that beta amplitude is normalized to the individual total-amplitude (1-30Hz) to account for unspecific differences (e.g., skull thickness) between participants.

World Health Organization Quality of Life Assessment (WHOQOL)

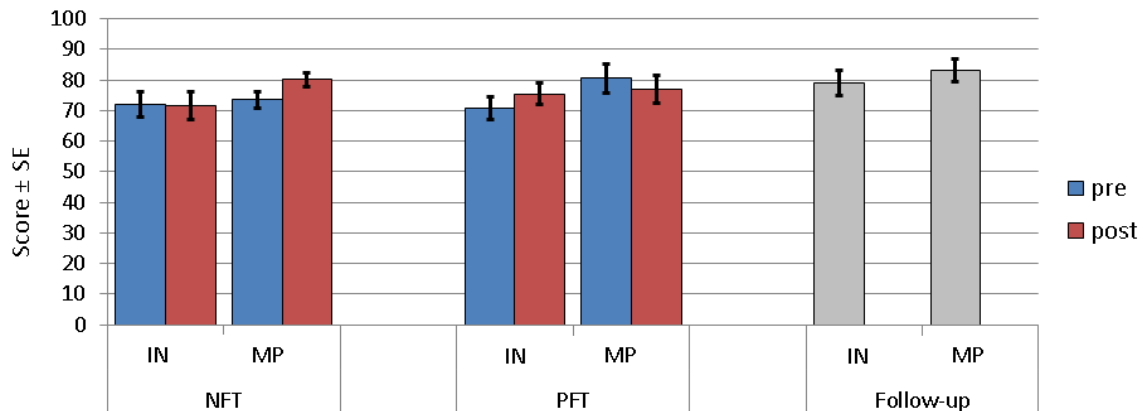
As for PSQI we ran an ANOVA with the factors PRE/POST (before or after 12 training blocks), FEEDBACK (Placebo- or Neurofeedback Training) and the between-subject factor GROUP (Insomnia, Insomnia Misperception) for the dependent measure subjective life quality (as assessed by the WHOQOL).

The ANOVA revealed no main effects for the factors FEEDBACK ($F_{1,21} = .001$, $p = .98$; $\eta_p^2 < .01$) and GROUP ($F_{1,21} = .002$, $p = .96$; $\eta_p^2 < .01$), but a main effect for the factor FEEDBACK ($F_{1,21} = 6.411$, $p = .02$; $\eta_p^2 = .234$). Yet, none of the interactions with FEEDBACK and PRE/POST were significant. The latter indicates a generally higher global life quality during the PFT condition yet no systematic change towards better quality of life after PFT and/or NFT training.

WHOQOL - GLOBAL

Supplementary Fig. 5. Subjective life quality (WHOQOL) on the global scale. Quality of life (WHOQOL global scale) remained unaffected by NFT/PFT training. Yet overall, quality of life was rated slightly higher in the PFT as compared to the NFT condition. Abbreviations: INs, insomniacs; MPs, misperception insomniacs.

Focusing on the World Health Organization Quality of Life Assessment (WHOQOL) sub-dimensions we were able to confirm the unspecific increase of physical quality of life (incorporating facets such as fatigue, physical discomfort or work capacity) from the first experimental polysomnography night 1 to the 3-month follow-up ($t_{22}=-3.531$, $p=.002$) as reported earlier (Schabus et al., 2014) (cf. Supplementary Fig.6).

WHOQOL - PHYSICAL

Supplementary Fig. 6. Subjective life quality on the physical sub-domain of the WHOQOL. Note an unspecific increase from pre-treatment (NFT or PFT) to the follow-up after 3 months in both sleep state misperception insomniacs (MPs) as well as insomniacs (INs). Facets incorporated in the physical health domain include activities of daily living, energy and fatigue, pain and discomfort, sleep and rest or work capacity.

Subjective awakening quality (SSS and MDBF)

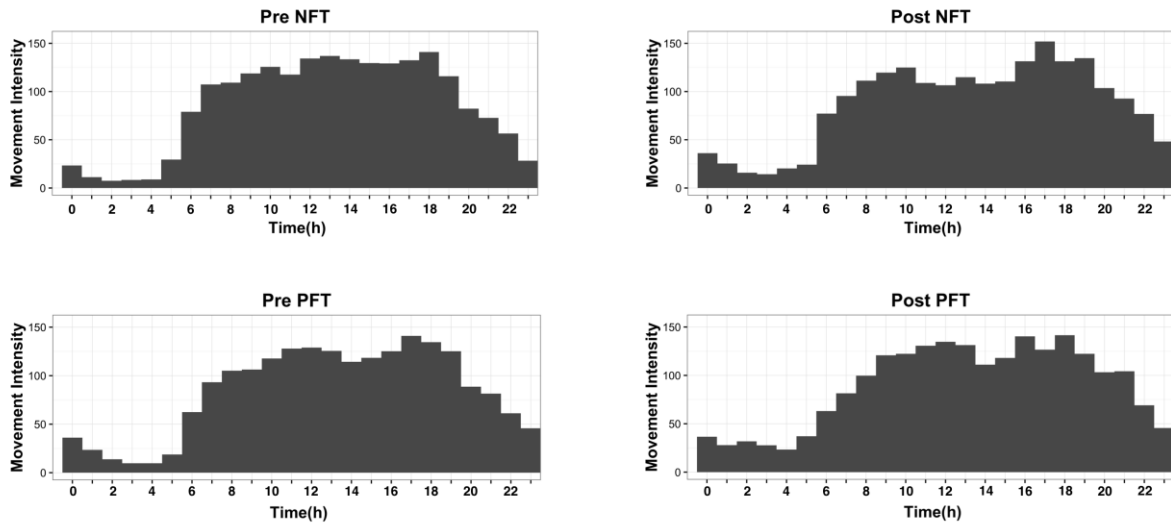
Stanford Sleepiness scales (Hoddes et al., 1972) (ranging from 1 full awake to 8 sleeping), generally indicate mild fatigue (across conditions) in the morning. Yet the change of the morning sleepiness was independent of whether participants received NFT or PFT training (PFT-first group: Pre to post PFT: $M=3.04$, $SD=0.72$ to $M=3.33$, $SD=0.89$; NFT-first group: Pre to post NFT: $M=3.62$, $SD=1.19$ to $M=3.00$, $SD=1.00$).

The Multidimensional Mood Questionnaire (MDBF; Steyer et al., 1997) quantifies the morning well-being on 3 sub-dimensions (good-bad mood [$M_{norm}=30$], alertness-tiredness [$M_{norm}=26$], and calmness-restlessness [$M_{norm}=28$] with high values indicating better well-being). Analyses indicate no change of any of the 3 sub-dimensions across NFT or PFT training (PFT-first group: Pre to post PFT [good-bad mood]: $M=29.75$, $SD=4.16$ to $M=30.71$, $SD=6.21$; Pre to post PFT [alertness-tiredness]: $M=21.15$, $SD=3.52$ to $M=23.60$, $SD=6.22$; Pre to post PFT [calmness-restlessness]: $M=30.33$, $SD=3.68$ to $M=30.75$, $SD=5.10$; NFT-first group: Pre to post NFT [good-bad mood]: $M=28.62$, $SD=6.59$ to $M=29.69$, $SD=6.97$; Pre to post NFT [alertness-tiredness]: $M=21.50$, $SD=5.73$ to $M=20.28$, $SD=7.05$; Pre to post NFT [calmness-restlessness]: $M=31.12$, $SD=5.47$ to $M=32.35$, $SD=5.18$).

Actigraphy Data

In addition to analyzing sleep data from questionnaires and objective polysomnographies we also acquired actigraphy throughout the experiment (starting at least 7 days before an experimental polysomnography night). Unfortunately not all patients wore the actigraph reliably on a daily basis which leaves us with only 17 for a pre-post comparison, before to after the NFT and PFT training blocks. Given the limited sample size we here pooled misperception insomniacs (MPs) and insomnia (IN) patients. The analysis checked for the average actigraphy activity across 7.25 hours (being the average total sleep time of our sex and age-matched healthy controls; Suppl. Table 1), and fixed the lights out time to the average of our patients (~22:30); only weekdays were considered in order to have a more reliable measure of nightly activity during normal working days across participants.

Two paired-sample t-tests revealed no significant differences in nightly activity from pre to post PFT blocks ($t_{16}=-0.72$, $p=0.48$) or from pre to post NFT blocks ($t_{17}=-1.24$, $p=0.23$). For an illustration of the averaged actigraphy data of our patients across the weeks preceding and following NFT/PFT training please refer to Supplementary Fig. 7.



Supplementary Fig. 7. Average actigraphy activity preceding and following NFT/PFT training blocks. Nightly activity was averaged across a week (excluding weekends) for 7.25hrs blocks around the average bed time (about 22:30). Note that patients (INs and MPs pooled) did not present less nightly activity (or less fragmented sleep) following NFT (or PFT).

The following table reflects sleep data computed from the mean of two experimental nights (finger-tapping and word-pair learning night) and following a screening/adaptation night for the healthy sleep control (SC) group.

Supplementary Table 1. Sleep data from sex and age-matched healthy controls.

HEALTHY CONTROLS	
(N=31)	
TIB (min)	477.8 (± 10.0)
TST (min)	436.6 (± 21.7)
Seff	91.4 (± 3.9)
SOL (min)	22.5 (± 14.2)
WASO (min)	24.4 (± 14.4)
NOA	15.5 (± 7.0)
N1 (%)	14.1 (± 5.9)
N2 (%)	46.8 (± 6.1)
N3 (%)	20.1 (± 8.6)
R (%)	19.0 (± 4.9)

NFT effects on memory

In order to test the effect of neurofeedback on memory we computed two ANOVAs (one for the procedural finger-tapping, and one for the declarative word-pair task) with the factors FEEDBACK (NFT vs. PFT) and PRE/POST (memory performance preceding and following the 12 NFT/PFT sessions). Dependent variables were the difference scores from the evening performance before falling asleep to the memory scores following a full (polysomnography) night and after an interference intervention (i.e., learning conflicting material such as new word-pairs or new finger-tapping sequences before a final recall of the originally learned material).

The ANOVA for the procedural task revealed no significant main effects for the factors FEEDBACK ($F_{1,19} = 1.10$; $p = .31$; $\eta_p^2 = .055$), PRE/POST ($F_{1,19} = 1.48$; $p = .24$; $\eta_p^2 = .072$) or their interaction ($F_{1,19} = 0.39$; $p = .85$; $\eta_p^2 = .002$). Likewise, the ANOVA for the declarative task revealed no significant main effects for the factors FEEDBACK ($F_{1,19} = 0.02$; $p = .90$; $\eta_p^2 = .001$), PRE/POST ($F_{1,19} = 2.08$; $p = .17$; $\eta_p^2 = .099$) or their interaction ($F_{1,19} = 2.03$; $p = .17$; $\eta_p^2 = .097$).

References

- Hauri P. Treating psychophysiologic insomnia with biofeedback. *Arch Gen Psychiatry* 1981; 38(7): 752-8.
- Hauri PJ, Percy L, Hellekson C, Hartmann E, Russ D. The treatment of psychophysiologic insomnia with biofeedback: a replication study. *Biofeedback Self Regul* 1982; 7(2): 223-35.
- Hoddes E, Zarcone V, Dement W. Development and Use of Stanford Sleepiness Scale (SSS). *Psychophysiology* 1972; 9: 150.
- Perlis ML, Smith MT, Andrews PJ, Orff H, Giles DE. Beta/Gamma EEG activity in patients with primary and secondary insomnia and good sleeper controls. *Sleep* 2001; 24(1): 110-7.
- Steyer R, Schwenkmezger P, Notz P, Eid M. The multidimensional mental state questionnaire (Mehrdimensionale Befindlichkeitsfragebogen (MDBF)): Manual. Göttingen, Germany: Hogrefe; 1997.